

**SVEUČILIŠTE U ZAGREBU**  
**PRIRODOSLOVNO–MATEMATIČKI FAKULTET**  
**MATEMATIČKI ODSJEK**

Mislav Đurkan

**STATISTIČKE METODE ZA**  
**ODREĐIVANJE ROKA TRAJANJA**  
**LIJEKOVA**

Diplomski rad

Voditelji rada:  
doc.dr.sc. Payle Goldstein i  
dr.sc. Drago Špoljarić

Zagreb, 2017.

Ovaj diplomski rad obranjen je dana \_\_\_\_\_ pred ispitnim povjerenstvom u sastavu:

1. \_\_\_\_\_, predsjednik
2. \_\_\_\_\_, član
3. \_\_\_\_\_, član

Povjerenstvo je rad ocijenilo ocjenom \_\_\_\_\_.

Potpisi članova povjerenstva:

1. \_\_\_\_\_
2. \_\_\_\_\_
3. \_\_\_\_\_

# Sadržaj

<b>Sadržaj</b>	<b>iii</b>
<b>Uvod</b>	<b>1</b>
<b>1 Pristup problemu</b>	<b>2</b>
1.1 Pozadina problema . . . . .	2
1.2 Regulatoriva . . . . .	2
1.3 Način testiranja . . . . .	3
<b>2 Analiza stabilnosti sa fiksnim serijama</b>	<b>5</b>
2.1 Preliminarni test za jednakost serija . . . . .	5
2.2 ANCOVA . . . . .	6
2.3 Minimalni pristup za višestruke serije . . . . .	10
<b>3 Praktična primjena</b>	<b>13</b>
3.1 Primjeri . . . . .	13
3.2 Problemi metode . . . . .	20
3.3 Implementacija u R-u . . . . .	24
<b>Bibliografija</b>	<b>29</b>

# Uvod

Za svaki lijek koji se stavlja na tržište, regulatorne agencije kao što su FDA u SAD-u ili EMA u EU zahtijevaju da se jasno označi njegov pripadni rok trajanja (eng. *shelf-life*). Rok trajanja daje potrošaču povjerenje da će lijek zadržati svoj identitet, snagu, kvalitetu i čistoću kroz navedeni rok. Da bi se to osiguralo provodi se studija stabilnosti za prikupljanje, analizu i tumačenje podataka o stabilnosti lijeka kroz određeni vremenski period. Statistika igra značajnu ulogu u dizajniranju i analizi studija stabilnosti kako bi se na pravi način procijenila degradacija određenog lijeka te kako bi se što preciznije procijenio njegov rok trajanja. U radu ću opisati metodu analize kovarijance i način na koji se ona primjenjuje u datom slučaju. Nakon toga opisat ću kako odrediti rok trajanja određenog lijeka. Na kraju ću dati par primjena navedenih metoda na stvarnim podacima i komentirati neke od njihovih nedostataka.

# Poglavlje 1

## Pristup problemu

### 1.1 Pozadina problema

Za neke lijekove, poput penicilina, znalo se da su nestabilni. Sve do početka 1970. nije bilo propisa vezanih uz stabilnost lijeka. Tada je nastala zabrinutost da nestabilni lijek neće moći održati svoj identitet, snagu, kvalitetu i čistoću nakon skladištenja u određenom vremenskom periodu. Kako bi se osigurala navedena svojstva, United States Pharmacopeia (USP) je 1975. uključila klauzulu o datumu isteka roka trajanja lijekova. Rok trajanja trebao bi osigurati da će u tom vremenskom periodu lijek imati sastav kako je i navedeno na samom pakiranju lijeka. U lijeku postoje dvije vrste sastojaka, aktivni i neaktivni. Aktivni sastojak definira se kao bilo koja komponenta lijeka koja ima izravan utjecaj u liječenju, ublažavanju boli ili prevenciji neke bolesti. Svi ostali sastojci nazivaju se neaktivnima. Zanima nas kako se aktivni sastojak mijenja s obzirom na vrijeme. Recimo da imamo Ibuprofen Belupo od 400mg, zanima nas do kada i s kojom sigurnošću možemo tvrditi da će u svakoj tableti udio ibuprofena biti jednak 400mg uz mala odstupanja. Ispituje se kakav utjecaj ima vrijeme na promatrani aktivni sastojak (ibuprofen) pri skladištenju u određenim uvjetima. United States Food and Drug Administration 1984. izdaje prvu regulativu za stabilnost lijekova. Godine 1993. International Conference on Harmonization (ICH) izdaje regulativu za stabilnost lijekova koja objedinjuje sve regulative do tada i čija svrha je definirati stvari koje su potrebne za prijavu registracije novog lijeka ili lijekovite tvari unutar Europske Unije, Japana i Amerike.

### 1.2 Regulativa

Stabilnost lijeka definirana je kao kapacitet lijeka da ostane unutar zadanih specifikacija kako bi se osigurao identitet, snaga, kvaliteta i čistoća lijeka. Serija lijeka definira se kao određena količina lijeka za koju je zamišljeno da ima jedinstven karakter i kakvoću, unutar

određenih granica, i proizvodi se u skladu s narudžbom od proizvođača tijekom istog ciklusa proizvodnje. Cilj studije stabilnosti je opisati degradaciju lijeka te utvrditi rok trajanja koji je primjenjiv na sve buduće serije tog lijeka. Nadalje, svrha studije je i pružiti dokaze o tome kako kvaliteta ljekovite tvari ili lijeka varira s obzirom na vrijeme uz utjecaj različitih prirodnih čimbenika poput temperature, vlažnosti i svjetlosti te uspostaviti period kada se ponovno treba provesti testiranje roka trajanja i uvjeta skladištenja. Da bi se provela studija stabilnosti, treba osigurati podatke za najmanje tri serije lijeka. Svrha toga je kvantificiranje varijabilnost između serija. Cilj je dozvoliti neku procjenu varijabilnosti između serija i testirati hipotezu da je opravdano koristiti jedan rok trajanja za sve serije. Serije bi trebale biti iste formulacije i u pakiranjima u kojima će ići na tržište. Proces proizvodnje korišten za te serije trebao bi simulirati onaj koji će biti primijenjen u proizvodnji za tržište, te bi trebao osigurati proizvod iste kvalitete i specifikacija kako je i zamišljeno za izlazak na tržište. Gdje je moguće, serije lijeka trebale bi biti proizvedene koristeći različite količine ljekovite tvari. Studiju bi trebalo primijeniti i na različite veličine pakiranja. Također, studija bi trebala uključivati testiranje onih karakteristika lijeka koje su podložne promjeni tijekom skladištenja i koje vjerojatno utječu na kvalitetu, sigurnost i efikasnost lijeka. Testiranje bi trebalo pokriti fizičke, kemijske i biološke karakteristike, sadržaj konzervansa i test funkcionalnosti (način uzimanja doze). Za proizvod s predloženim rokom trajanja od najmanje 12 mjeseci, frekvencija ponovljenog testiranja trebala bi biti svaka 3 mjeseca kroz prvu godinu, svakih 6 mjeseci kroz drugu godinu i analogno dalje kroz cijeli period predloženog roka trajanja.

### 1.3 Način testiranja

Recimo da smo proveli studiju u kojoj smo utvrdili kako se promatrani sastojci ponašaju s obzirom na vrijeme, skladištenjem u dogovorenim uvjetima. Fokus našeg ispitivanja bit će aktivni sastojak lijeka. Kao što smo spomenuli, testiranje se provodi s minimalno tri različite serije lijeka. Za svaku seriju možemo promatrati njen degradacijski pravac, tj. kako se udio promatranog sastojka mijenja (smanjuje) s obzirom na vrijeme. Cilj je na neki način testirati sličnost promatranih serija te opravdati korištenje jednog roka trajanja za sve serije zajedno. Regulativa nalaže da se sličnost serija, tj. pravaca degradacije za serije može promatrati u smislu jednakosti odsječka i jednakosti nagiba dobivenih iz podataka za svaku seriju zasebno. Ako je varijabilnost između serija mala, bilo bi povoljno gledati sve podatke kao jednu seriju te tako procijeniti jedinstveni rok trajanja s većom preciznošću. Testiranje sličnosti serija provest ćemo koristeći metodu analize kovarijance (ANCOVA). ANCOVA je u suštini kombinacija metode analize varijance (ANOVA) i regresijske analize. Jednostavna linearna regresija je statistička tehnika koja definira odnos između dvije varijable ( $X$  i  $Y$ ) pomoću pravca koji ih najbolje opisuje, u smislu metode najmanjih kvadrata. Pravac je opisan jednadžbom  $Y=A+BX$  gdje je  $Y$  zavisna,  $X$  nezavisna varijabla,  $A$  je odsječak

na y-osi i B je nagib pravca. ANOVA se koristi kako bi se utvrdilo razlikuju li se srednje vrijednosti tri ili više grupa podataka (označenih nekom faktorskom varijablom) baziranih na neprekidnoj zavisnoj varijabli, koju nazivamo varijabla odziva. Faktorska varijabla nam služi kako bi razlikovali pripadnost podataka pojedinoj grupi. ANCOVA se razlikuje po tome što uspoređuje varijablu odziva s obzirom na faktorsku varijablu i s obzirom na neprekidnu nezavisnu varijablu, koju nazivamo kovarijatom. Kovarijate su varijable koju ne želimo ispitivati u studiji. Točnije, to su varijable koje predstavljaju izvor varijabilnosti za koji se smatra da utječe na zavisnu varijablu, ali nisu bile kontrolirane u eksperimentalnim procedurama. Ovisno o rezultatima testiranja sličnosti serija, rok trajanja lijeka računamo zasebno za svaku seriju ili (u idealnom slučaju) za sve serije zajedno. Regulativa nalaže da je rok trajanja vrijeme u kojem donja granica pouzdanog intervala očekivanog pravca degradacije, koji je dobiven metodom najmanjih kvadrata, siječe određenu specifikacijsku granicu. Specifikacijske granice, ako nisu propisane farmakopejom, obično se definiraju tijekom razvoja lijeka koristeći farmakokinetika saznanja te rezultate preliminarnih testiranja. Primjerice, za sadržaj lijeka je granica propisana farmakopejom te iznosi od 90% do 110% propisanog sadržaja na pakiranju.

## Poglavlje 2

# Analiza stabilnosti sa fiksnim serijama

### 2.1 Preliminarni test za jednakost serija

U ovom poglavlju koristimo rezultate iz [2]. Kao što smo spomenuli, testirat ćemo sličnosti promatranih serija u smislu jednakosti odsječka i jednakosti nagiba dobivenih iz podataka za svaku seriju zasebno. Neka je  $y_{ij}$  postotak aktivnog sastojka, tj. postotak sadržaja lijeka za koji proizvođač tvrdi da je prisutan za  $i$ -tu seriju lijeka u vremenu uzorkovanja  $x_{ij}$ . Ako sadržaj opada linearno kroz vrijeme, za sve serije, model koji opisuje degradaciju za više serija lijekova izgleda ovako:

$$y_{ij} = \alpha_i + \beta_i x_{ij} + e_{ij}, \quad i = 1, \dots, K, \quad j = 1, \dots, n_i, \quad (2.1)$$

gdje su  $\alpha_i$  i  $\beta_i$  odsječak i nagib pravca degradacije za seriju  $i$ , dok za  $e_{ij}$  pretpostavljamo da su nezavisne, jednako distribuirane, normalne slučajne varijable s očekivanjem 0 i varijansom  $\sigma^2$ . Primijetimo da  $\alpha_i$  možemo promatrati kao efekt  $i$ -te serije u vremenu 0, tj. sadržaj na početku studije stabilnosti, a  $\beta_i$  kao stopu degradacije u odnosu na vrijeme. Testiranje hipoteza za jednakost odsječka i nagiba je u suštini test homogenosti pravca degradacije između serija, koji može biti ispitan testiranjem navedenih hipoteza:

$$H_{0\alpha} : \alpha_i = \alpha_{i'} \quad \text{za sve } i \neq i' \quad (2.2)$$

$$H_{0\beta} : \beta_i = \beta_{i'} \quad \text{za sve } i \neq i' \quad (2.3)$$

Testiranje hipoteza (2.2) i (2.3) može se provesti koristeći analizu kovarijance (ANCOVA). Također, regulativa nalaže da se testiranje obavi na razini značajnosti od 25%. U [1] je obrazloženo da je definirana veća razina značajnosti kako bi se povećala šansa da se odbaci hipoteza o sličnosti nagiba tj. kako bi teže izračunali duži rok trajanja kao posljedicu ujedinjenja podataka. S jedne strane, ovo se može činiti kao konzervativno pravilo kojim



želimo osigurati da ne precjenjujemo rok trajanja lijeka. Međutim, takva odluka naišla je na mnoge kritike. U primjeru 3.2 obrazložiti ćemo problem nastao zbog izbora navedene razine značajnosti.

## 2.2 ANCOVA

U ovome poglavlju opisat ću metodu analize kovarijance u terminima koji su vezani za opisanu problematiku. Promatrane varijable su vrijeme (kovarijata), serija (faktorska varijabla) i postotak sadržaja lijeka (varijabla odziva). Neka su  $S_{xx}(i)$ ,  $S_{yy}(i)$ ,  $S_{xy}(i)$  sume kvadrata vremena, postotka sadržaja lijeka i njihovog skalarnog produkta za svaku seriju  $i$  posebno:

$$\begin{aligned} S_{xx}(i) &= \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 \\ S_{yy}(i) &= \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \\ S_{xy}(i) &= \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i) \end{aligned} \quad (2.4)$$

i neka je

$$\begin{aligned} S_{xx}(W) &= \sum_{i=1}^K S_{xx}(i) \\ S_{yy}(W) &= \sum_{i=1}^K S_{yy}(i) \\ S_{xy}(W) &= \sum_{i=1}^K S_{xy}(i), \end{aligned} \quad (2.5)$$

gdje  $K$  označuje koliko imamo serija,  $n_i$  je broj opažanja u seriji  $i$ ,  $\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$  i  $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$ . U [4] je pokazano da su procjene nagiba i odsječka degradacijskog pravca metodom najmanjih kvadrata dane sa:

$$\hat{\beta}_i = \frac{S_{xy}(i)}{S_{xx}(i)}$$

i

$$\hat{\alpha}_i = \bar{y}_i - \hat{\beta}_i \bar{x}_i.$$

Suma kvadrata reziduala za svaku seriju  $i$  definira se kao:  $SSE(i) = \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_i)^2$ , gdje je  $\hat{y}_i = \hat{\alpha}_i + \hat{\beta}_i x_{ij}$ . U [4] je pokazano da vrijedi:

$$SSE(i) = S_{yy}(i) - \frac{[S_{xy}(i)]^2}{S_{xx}(i)}. \quad (2.6)$$

Kombinirana suma kvadrata reziduala je:

$$SSE = \sum_{i=1}^K SSE(i),$$

sa  $N - 2K$  stupnjeva slobode. SSE je najmanja moguća greška dobivena kao zbroj grešaka za svaku seriju zasebno. Nema pretpostavki vezanih uz nagib i odsječak pravca. Suma kvadrata reziduala dobivena iz  $S_{xx}(W)$ ,  $S_{yy}(W)$  i  $S_{xy}(W)$  dana je sa:

$$SSE(W) = S_{yy}(W) - \frac{[S_{xy}(W)]^2}{S_{xx}(W)}, \quad (2.7)$$

sa  $N - K - 1$  stupnjeva slobode. To je greška uz pretpostavku da pravci imaju jednak nagib, ali potencijalno različite odsječke. Procjena zajedničkog nagiba, metodom najmanjih kvadrata dana je sa:

$$\hat{\beta} = \frac{S_{xy}(W)}{S_{xx}(W)}. \quad (2.8)$$

Pokažimo navedenu tvrdnju. Želimo pronaći točku u kojoj funkcija:

$$L(\alpha_i, \beta) = \sum_{i=1}^K \sum_{j=1}^{n_i} (\alpha_i + \beta x_{ij} - y_{ij})^2$$

ima globalni minimum. Deriviranjem po  $\alpha_i$  i  $\beta$  dobijemo:

$$\frac{\partial L}{\partial \alpha_i} = 2 \sum_{j=1}^{n_i} (\alpha_i + \beta x_{ij} - y_{ij}) = 0, \quad i = 1, \dots, K$$

i

$$\frac{\partial L}{\partial \beta} = 2 \sum_{i=1}^K \sum_{j=1}^{n_i} (\alpha_i + \beta x_{ij} - y_{ij}) x_{ij} = 0.$$

Odnosno:

$$\alpha_i n_i + \beta n_i \bar{x}_i - n_i \bar{y}_i = 0 \quad (2.9)$$

i

$$\sum_{i=1}^K \alpha_i n_i \bar{x}_i + \beta \sum_{i=1}^K \sum_{j=1}^{n_i} x_{ij}^2 - \sum_{i=1}^K \sum_{j=1}^{n_i} x_{ij} y_{ij} = 0 \quad (2.10)$$

Sada raskisujemo  $S_{xx}(W)$  kao:

$$S_{xx}(W) = \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 = \sum_{i=1}^K \sum_{j=1}^{n_i} x_{ij}^2 - \sum_{i=1}^K \bar{x}_i^2 n_i. \quad (2.11)$$

Analogno:

$$S_{xy}(W) = \sum_{i=1}^K \sum_{j=1}^{n_i} x_{ij} y_{ij} - \sum_{i=1}^K \bar{x}_i \bar{y}_i n_i. \quad (2.12)$$

Pomnožimo (2.9) sa  $\bar{x}_i$  i izvučemo van  $\alpha_i n_i \bar{x}_i$ . Uvrstimo u (2.10) i koristeći (2.11) i (2.12) dobijemo (2.8). Sada pogledajmo je li naša pretpostavka o jednakosti nagiba doprinijela poboljšanju modela. Suma kvadrata za razlike u nagibu dana je sa:

$$SS(\beta) = SSE(W) - SSE.$$

Prema tome, nultu hipotezu o jednakosti nagiba odbacujemo na razini značajnosti  $\alpha$  ako:

$$F_\beta = \frac{MS(\beta)}{MSE} > F_{\alpha, K-1, N-2K},$$

gdje je

$$\begin{aligned} MS(\beta) &= \frac{SS(\beta)}{K-1} \\ MSE &= \frac{SSE}{N-2K} \end{aligned} \quad (2.13)$$

i  $F_{\alpha, K-1, N-2K}$  je  $\alpha$ -kvantil centralne F distribucije sa  $K-1$  i  $N-2K$  stupnjeva slobode. Ako  $SSE(W)$  značajno poraste u odnosu na  $SSE$ ,  $F_\beta$  naraste, a time se smanjuje p-vrijednost za hipotezu o jednakosti nagiba, tj. pretpostavkom jednakosti nagiba nismo poboljšali naš model te ju odbacujemo na razini značajnosti  $\alpha$ . Ako ne odbacujemo nultu hipotezu o jednakosti nagiba na razini značajnosti  $\alpha$ ,  $\beta_i$  u modelu (2.1) može se zamjeniti zajedničkim nagibom  $\beta$ . Tada se model (2.1) reducira na:

$$y_{ij} = \alpha_i + \beta x_{ij} + e_{ij}, \quad i = 1, \dots, K, \quad j = 1, \dots, n_i. \quad (2.14)$$

Sada definiramo:

$$\begin{aligned}
 S_{xx}(T) &= \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{..})^2 \\
 S_{yy}(T) &= \sum_{i=1}^K \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2 \\
 S_{xy}(T) &= \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{..})(y_{ij} - \bar{y}_{..})
 \end{aligned}
 \tag{2.15}$$

$$\begin{aligned}
 S_{xx}(B) &= \sum_{i=1}^K n_i (\bar{x}_i - \bar{x}_{..})^2 \\
 S_{yy}(B) &= \sum_{i=1}^K n_i (\bar{y}_i - \bar{y}_{..})^2
 \end{aligned}$$

$$S_{xy}(B) = \sum_{i=1}^K n_i (\bar{x}_i - \bar{x}_{..})(\bar{y}_i - \bar{y}_{..}),$$

gdje je  $N = \sum n_i$  i

$$\begin{aligned}
 \bar{x}_{..} &= \frac{1}{N} \sum_{i=1}^K \sum_{j=1}^{n_i} x_{ij} \\
 \bar{y}_{..} &= \frac{1}{N} \sum_{i=1}^K \sum_{j=1}^{n_i} y_{ij}.
 \end{aligned}$$

Analogno kao u (2.6), ukupna suma devijacija od regresije dana je sa:

$$SST = S_{yy}(T) - \frac{[S_{xy}(T)]^2}{S_{xx}(T)}.$$

To je ukupna greška uz pretpostavku da su nagib i odsječak jednaki za sve serije, tj. da imamo jedinstven pravac. Sada pogledajmo je li naša pretpostavka o jednakosti nagiba i odsječka doprinijela poboljšanju modela u odnosu na model koji ima jednake nagibe, a različite odsječke. Gledamo sljedeću sumu:

$$SSB = SST - SSE(W).$$

Stupnjevi slobode za  $SST$ ,  $SSW$  i  $SSB$  su redom  $N - 2$ ,  $N - K - 1$  i  $K - 1$ . Prema tome, imamo:

$$MSB = \frac{SSB}{K - 1}$$

i MSE definiran kao u (2.13). Nultu hipotezu za jednakost odsječaka odbacujemo ako vrijedi:

$$F_{\alpha} = \frac{MSB}{MSE} > F_{\alpha, K-1, N-2K},$$

gdje je  $F_{\alpha, K-1, N-2K}$   $\alpha$ -kvantil centralne F distribucije sa  $K - 1$  i  $N - 2K$  stupnjeva slobode. Suma kvadrata s obzirom na zajednički nagib dana je sa:

$$SSS = \frac{[S_{xy}(W)]^2}{S_{xx}(W)}.$$

Rezultate zapišimo u sljedeću tablicu:

Izvor varijabilnosti	Broj stupnjeva slobode	Suma kvadrata odstupanja	Srednje kvadratno odstupanje	F-statistika
Odsječak (serija)	K-1	SSB	MSB	$F_{\alpha} = \frac{MSB}{MSE}$
Vrijeme (zajednički nagib)	1	SSS	MSS	$F_s = \frac{MSS}{MSE}$
Razlika u nagibu	K-1	$SS(\beta)$	$MS(\beta)$	$F_{\beta} = \frac{MS(\beta)}{MSE}$
Greška	N-2K	SSE	MSE	
Ukupno	N-1	SST		

Tablica 2.1: ANCOVA za model (2.1)

## 2.3 Minimalni pristup za višestruke serije

### 1. način

Ako preliminarnim testom zaključimo da ne odbacujemo hipoteze (2.2) i (2.3) na razini značajnosti od 25%, smatramo da su sve serije iz iste populacije neke proizvodne serije sa zajedničkim degradacijskim uzorkom. Model (2.1) reducira se na:

$$y_{ij} = \alpha + \beta x_{ij} + e_{ij}, \quad i = 1, \dots, K, \quad j = 1, \dots, n_i \quad (2.16)$$

gdje su  $\alpha$  i  $\beta$  zajednički nagib i odsječak te vrijedi ista pretpostavka normalnosti za  $e_{ij}$ . Neka je  $(\hat{\alpha}_C, \hat{\beta}_C)$  procjena od  $(\alpha, \beta)$  dobivena na podacima  $(x_{ij}, y_{ij})$  modela (2.16), odnosno:

$$\hat{\beta}_C = \frac{S_{xy}(T)}{S_{xx}(T)}$$

$$\hat{\alpha}_C = \bar{y}_{..} - \hat{\beta}_C \bar{x}_{..}$$

Za svako fiksno vrijeme  $x$ , donja granica 95%-tnog jednostranog intervala pouzdanosti za  $\alpha + \beta x$  dana je sa:

$$L(x) = \hat{\alpha} + \hat{\beta}x - \hat{\sigma} t_{n-2} \sqrt{\frac{1}{N} + \frac{(x - \bar{x}_{..})^2}{S_{xx}(T)}} \quad (2.17)$$

gdje je  $t_{n-2}$  95-ti percentil t-distribucije sa  $N - 2$  stupnjeva slobode,

$$\hat{\sigma}^2 = \frac{S_{yy}(T) - \hat{\beta}_C S_{xy}(T)}{N - 2}$$

i gdje su  $S_{xx}(T)$ ,  $S_{yy}(T)$ ,  $S_{xy}(T)$ ,  $\bar{x}_{..}$ ,  $\bar{y}_{..}$  definirani kao u (2.15). Dokaz tvrdnje (2.17) može se pronaći u [4]. Procjena očekivane vrijednosti degradacijskog pravca izračunatog metodom najmanjih kvadrata u vremenu  $t = x$  je:

$$y(x) = \hat{\alpha}_C + \hat{\beta}_C x,$$

pri čemu je procjenitelj za varijancu dan s:

$$\hat{V}[y(x)] = \hat{\sigma}^2 \left[ \frac{1}{N} + \frac{(\bar{x} - \bar{x}_{..})^2}{S_{xx}(T)} \right].$$

Kako bi se odredila procjena shelf-lifea, FDA predlaže sljedeću procjenu:

$$\hat{\theta}_F = \inf\{x \geq 0 : L(x) \leq \eta\},$$

gdje je  $\eta$  zadana specifikacijska granica. Prema tome, točke gdje jednadžba (2.17) siječe prihvatljivu specifikacijsku granicu su dva korijena sljedeće jednadžbe:

$$[\eta - (\hat{\alpha} + \hat{\beta}x)]^2 = \hat{\sigma}^2 t_{n-2}^2 \left[ \frac{1}{N} + \frac{(x - \bar{x}_{..})^2}{S_{xx}(T)} \right]. \quad (2.18)$$

Dva korijena (označimo ih sa  $x_L$  i  $x_U$ ) jednadžbe (2.18) čine gornji i donji limit za dvostrani 90% pouzdani interval od  $(\eta - \alpha)/\beta$ . Označimo:

$$SE(x) = \hat{\sigma}^2 \left[ \frac{1}{N} + \frac{(x - \bar{x}_{..})^2}{S_{xx}(T)} \right]^{\frac{1}{2}}$$

i neka je:

$$T_{\hat{\alpha}} = \frac{\hat{\alpha} - \eta}{SE(\hat{\alpha})} \quad \text{i} \quad T_{\hat{\beta}} = \frac{\hat{\beta}}{SE(\hat{\beta})}.$$

Ako je nagib statistički značajno manji od 0 i odsječak statistički značajno veći od  $\eta$  (što je prihvatljivi donji specifikacijski limit za 5% značajnosti) tj. ako je:

$$(a) \quad T_{\hat{\beta}} < -t_{0.05, n-2} \quad (2.19)$$

$$(b) \quad T_{\hat{\alpha}} > t_{0.05, n-2} \quad (2.20)$$

tada je 90% pouzdani interval za  $(\eta - \alpha)/\beta$  uključiv i zatvoren interval  $[x_L, x_U]$ . U tom slučaju rok trajanja definira se kao  $\hat{\theta}_F = x_L$ . Međutim, u ostalim slučajevima 90% pouzdani interval za  $(\eta - \alpha)/\beta$  je ili cijela realna linija ili dva odvojena otvorena intervala. Posljedično, tu nije definiran rok trajanja.

## 2. način

Ako preliminarnim testom zaključimo da odbacujemo nultu hipotezu o sličnosti degradacijskog pravca unutar serija, pravce ne možemo smatrati istima zbog razlike u nagibu i odsječku. U tom slučaju, prema FDA regulativi, cjelokupni rok trajanja ovisi o minimalnom vremenu za koji se očekuje da će serija ostati unutar dozvoljenih granica. Prema tome, neka je  $x_L(i)$  procijenjeni rok trajanja za seriju  $i$ ,  $i = 1, \dots, K$ . Cjelokupni rok trajanja je:

$$x_L(\min) = \min\{x_L(1), \dots, x_L(K)\}.$$

Kako je  $x_L(\min)$  najkraći rok trajanja među serijama, procjena će pružiti 95%-tnu pouzdanost da će lijek ostati iznad zadanog specifikacijskog limita  $\eta$  do  $x_L(\min)$  vremena, za sve serije. Međutim,  $x_L(\min)$  je konzervativna procjena cjelokupnog roka trajanja zato što daje više od 95%-tne pouzdanosti za sve serije, osim one za koju je procijenjeno. Takav pristup naziva se minimalni pristup.

## Poglavlje 3

# Praktična primjena

### 3.1 Primjeri

#### Primjer 1

Kako bi ilustrirali statističke metode koje smo opisali u poglavlju 2, promotrimo sljedeći primjer. Provedena je studija stabilnosti na tabletama od 300 mg da bi se odredio cjelokupni rok trajanja. Tablete iz 5 serija skladištene su na sobnoj temperaturi, a vrsta spremnika je boca. Testirana je snaga tableta svaka 3 mjeseca, zaključno s 15. mjesecom te snaga nakon 18 mjeseci. Postotak sadržaja lijeka za svaku seriju u određenom vremenu zapisan je u tablici 3.1. Nadalje, ispitivanje je provedeno u isto vrijeme za svaku seriju, što je najjednostavnija struktura studije stabilnosti.

Pakiranje	Serija	Vrijeme uzorkovanja					
		0	3	6	9	12	18
Boca	1	104.8	102.5	101.5	102.4	99.4	96.5
	2	103.9	101.9	103.2	99.6	100.2	98.8
	3	103.5	102.1	101.9	100.3	99.2	101.0
	4	101.5	100.3	101.1	100.6	100.7	98.4
	5	106.1	104.3	101.5	101.1	99.4	98.2

Tablica 3.1: Rezultati ispitivanja

Kako bi ilustrirali računanje roka trajanja za određenu seriju uzmimo podatke za seriju 1 iz Tablice 3.1. Računamo sljedeće:

$$\bar{x} = 8$$

$$\bar{y} = 101.183$$



$$\begin{aligned}S_{xx} &= 210 \\S_{xy} &= -88.9 \\S_{yy} &= 41.508.\end{aligned}$$

Nadalje,

$$\begin{aligned}\hat{\beta} &= \frac{S_{xy}}{S_{xx}} = \frac{-88.9}{210} = -0.423 \\ \hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x} = 101.183 - (-0.423)(8) = 104.57 \\ s^2 &= \frac{S_{yy} - \hat{\beta}S_{xy}}{n - 2} = \frac{41.508 - (-0.423)(-88.9)}{6 - 2} = 0.969\end{aligned}$$

Prema tome, očekivani degradacijski pravac procijenjen metodom najmanjih kvadrata u vremenu  $t = x$  dan je s:

$$y(x) = 104.57 - 0.423x.$$

Standardne greške procjenitelja za nagib i odsječak dane su s:

$$SE(\hat{\beta}) = \left[ \frac{s^2}{S_{xx}} \right]^{\frac{1}{2}} = \left[ \frac{0.969}{210} \right]^{\frac{1}{2}} = 0.0679$$

i

$$SE(\hat{\alpha}) = \left[ s^2 \left( \frac{1}{nS_{xx}} \sum_{j=1}^n x_j^2 \right) \right]^{\frac{1}{2}} = \left[ 0.969 \left( \frac{594}{6 * 210} \right) \right]^{\frac{1}{2}} = 0.676$$

Prije nego što primijenimo jednadžbu (2.18) kako bi izračunali rok trajanja, provjerimo (2.19) i (2.20). Za zadanu specifikacijsku granicu  $\eta = 90$

$$T_{\hat{\beta}} = T_{\hat{\beta}} = \frac{\hat{\beta}}{SE(\hat{\beta})} = \frac{-0.423}{0.0679} = -6.234$$

s  $p$  vrijednošću manjom od 0.001 i

$$T_{\hat{\alpha}} = \frac{\hat{\alpha} - \eta}{SE(\hat{\alpha})} = \frac{104.57 - 90}{0.676} = 21.55$$

s  $p$  vrijednošću od 0.0017. Zadovoljena su oba uvjeta. Prema tome, rok trajanja je procijenjen kao manji korijen jednadžbe:

$$[90 - (104.57 + (-0.423)x)]^2 = 2.132^2 * 0.969 \left[ \frac{1}{6} + \frac{(x - 8)^2}{210} \right],$$

što je 27.5. Dakle, procijenjeni rok trajanja za seriju 1 je 27.5 mjeseci. Jednadžbe (2.19) i (2.20) zadovoljene za sve serije osim serije 3 gdje je  $T_{\hat{\beta}} = 2.1$ , a  $p$  vrijednost je 0.053.

Dakle, nagib za seriju 3 nije značajno manji od 0 na razini značajnosti od 5%. Prema tome, rok trajanja za tu seriju ne može biti izračunat opisanim metodama. Učitajmo podatke u R i pokrenimo funkciju za računanje roka trajanja sa zadanom specifikacijskom granicom  $\eta = 90$  (implementacija funkcije može se pogledati u poglavlju 3.3) :

```
source("Shelf_life_funkcija.R")
bottle<-read_excel("bottle.xlsx")
shelf_life(bottle,90)
```

Dobijemo sljedeću tablicu:

Serija	Procijenjeni rok trajanja
1	27.46
2	33.45
3	-
4	51.43
5	28.36
Sve	39.6

Tablica 3.2: Rok trajanja izračunat iz podataka studije stabilnosti danih u tablici 3.1

Pokrenimo sada funkciju za računanje ANCOVA tablice dane u 2.1 (implementacija funkcije može se pogledati u poglavlju 3.3):

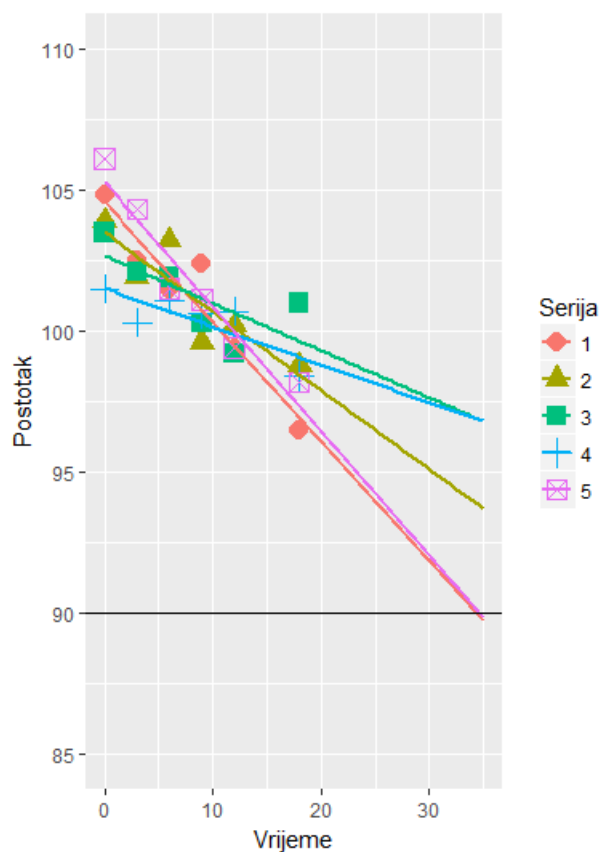
```
source("ANCOVA_funkcija.R")
ancova(bottle)
```

Izvor varijabilnosti	Stupanj slobode	Suma kvadrata	F	p
Odsječak	4	5.5879	1.4557	0.2528
Vrijeme	1	87.8416	91.5344	6.62e-09
Razlika u nagibu	4	16.7469	4.3627	0.0107
Greška	20	19.1931		

Tablica 3.3: ANCOVA za podatke iz tablice 3.1

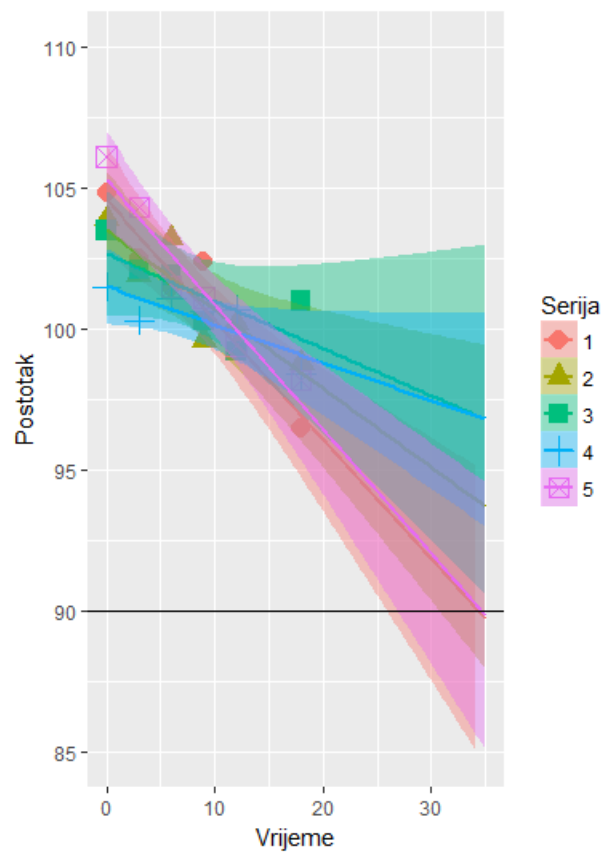
Vidimo da je  $p$  vrijednost za razliku u nagibu jednaka 0.0107 te na razini značajnosti od 25% odbacujemo hipotezu o jednakosti nagiba. Prema tome, kao što je opisano u poglavlju 2, procijenjeni rok trajanja je minimalni rok trajanja po serijama, što je u ovom slučaju 27.5 mjeseci.

Prikažimo sada metodu grafički. Prvo prikažimo dobivene podatke i pripadne prosječne pravce degradacije dobivene metodom najmanjih kvadrata za svaku seriju zasebno:



Slika 3.1: Prosječni pravci za podatke iz tablice 3.1

Vidimo da i vizualno možemo potkrijepiti našu tvrdnju da odbacujemo hipotezu o jednakosti nagiba. Prikažimo sada pouzdane intervale za svaki pravac i pogledajmo gdje se donja granica intervala siječe sa zadanom specifikacijskom granicom:



Slika 3.2: Pouzdani intervali oko pravaca regresije za svaku seriju zasebno

Vidimo da donja granica pouzdanog intervala za seriju 1 prva siječe zadanu specifikacijsku granicu u vremenu koje odgovara procijenjenom roku trajanja od 27.5 mjeseci.

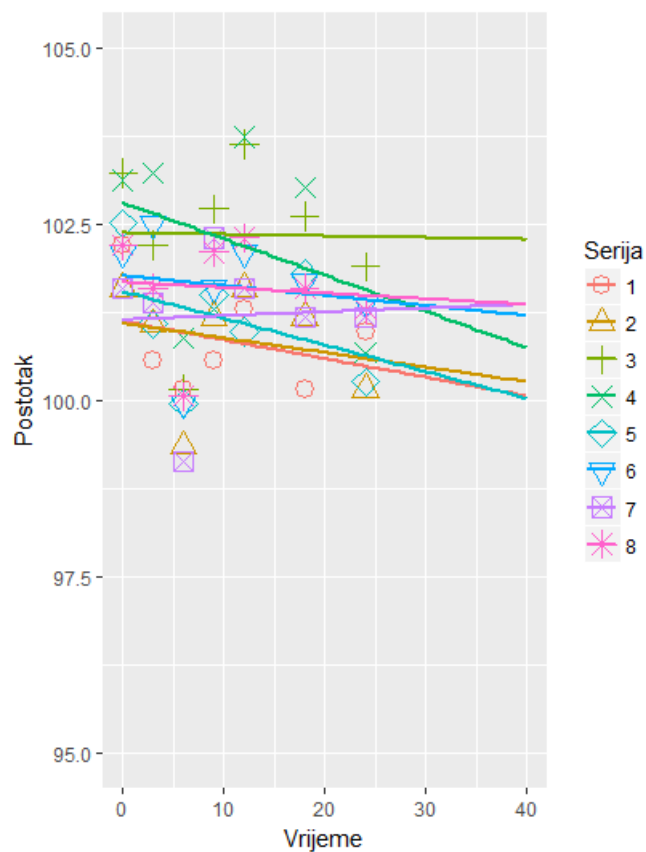
## Primjer 2

Dobiveni su sljedeći podaci kao rezultat provedbe studije stabilnosti:

Seriya	Vrijeme (mjesec)	Sadržaj (%)	Seriya	Vrijeme (mjesec)	Sadržaj (%)
1	0	102.206	5	0	102.512
	3	100.574		3	101.084
	6	100.166		6	99.962
	9	100.574		9	101.492
	12	101.288		12	100.982
	18	100.166		18	101.798
	24	100.982		24	100.268
2	0	101.594	6	0	102.104
	3	101.084		3	102.512
	6	99.35		6	99.962
	9	101.186		9	101.594
	12	101.594		12	102.104
	18	101.186		18	101.696
	24	100.166		24	101.39
3	0	103.226	7	0	101.594
	3	102.206		3	101.39
	6	100.166		6	99.146
	9	102.716		9	102.308
	12	103.634		12	101.594
	18	102.614		18	101.186
	24	101.9		24	101.186
4	0	103.124	8	0	102.206
	3	103.226		3	101.594
	6	100.88		6	100.064
	9	101.288		9	102.104
	12	103.736		12	102.308
	18	103.022		18	101.594
	24	100.676		24	101.288

Tablica 3.4: Rezultati studije stabilnosti

Vizualizirajmo dobivene podatke:



Slika 3.3: Pravci regresije za svaku seriju zasebno

Analogno kao u Primjeru 1 izračunamo tablicu za ANCOVU:

Izvor varijabilnosti	Stupanj slobode	Suma kvadrata	F	p
Odsječak	7	16.581	2.349	0.042
Vrijeme	1	1.295	1.284	0.264
Razlika u nagibu	7	1.067	0.151	0.993
Greška	40	40.339		

Tablica 3.5: ANCOVA za podatke iz tablice 3.4

Vidimo da je  $p$  vrijednost za razliku u nagibu jednaka 0.993 te na razini značajnosti od 25% ne možemo odbaciti hipotezu o jednakosti nagiba. Nadalje, vidimo da je  $p$  vrijednost za odsječak jednaka 0.042 te na razini značajnosti od 25% odbacujemo hipotezu o jednakosti odsječka. Prema tome, kako bi procijenili rok trajanja lijeka pretpostavljamo da model ima zajednički nagib i različite odsječke. Procjenitelj za nagib dan je sa (2.8). Rok trajanja za svaku seriju zasebno (uz navedene pretpostavke) izgleda ovako:

Serija	Procijenjeni rok trajanja
1	254.492
2	255.148
3	288.288
4	286.648
5	261.383
6	271.883
7	262.367
8	271.227

Tablica 3.6: Rok trajanja izračunat iz podataka studije stabilnosti danih u tablici 3.4

Dakle, kao što je opisano u poglavlju 2, procijenjeni rok trajanja je minimalni rok trajanja po serijama, što je u ovom slučaju 254.5 mjeseci.

## 3.2 Problemi metode

Procedure za testiranje sličnosti serija i minimalni pristup za određivanje roka trajanja primile su mnogo kritika zbog svojih nedostataka. Stephen J. Ruberg i James W. Stegeman konstruirali su primjer na kojem su ilustrirali nedostatke korištenja minimalnog pristupa:

Seriya	Vrijeme(godina)	Sadržaj(%)	Seriya	Vrijeme(godina)	Sadržaj(%)
1	0,014	100,4	4	0,066	100,4
	0,280	100,3		0,343	100,0
	0,514	99,7		0,533	99,5
	0,769	99,2		0,802	99,3
	1,074	98,9		1,033	99,3
	1,533	98,2		1,538	98,2
	2,030	97,3	5	0,011	100,5
	3,071	95,7		0,310	99,8
	4,049	94,5		0,624	99,1
2	0,022	100,7		1,063	98,4
	0,118	100,6	6	0,011	100,1
	0,272	100,3		0,310	99,5
	0,566	99,9		0,624	98,5
	1,165	98,6		1,063	98,4
	2,022	97,6			
	3,077	96,4			
3	0,025	100,2			
	0,275	99,7			
	0,547	99,2			
	0,797	99,0			
	1,041	98,8			
	2,058	96,4			
	2,519	96,2			

Tablica 3.7: Prvi skup podataka iz [3]

U ovome primjeru koncentrirat ćemo se na analizu jednakosti nagiba, kao najkritičnijem djelu ispitivanja sličnosti serija. Pripadna ANCOVA tablica za podatke navedene u tablici 3.7 izgleda ovako:

Izvor varijabilnosti	Stupanj slobode	Suma kvadrata	F	p
Odsječak	5	2.829	13.232	2.342e-06
Vrijeme	1	71.092	1662.399	2.166e-24
Razlika u nagibu	5	0.351	1.640	0.186
Greška	25	1.069		

Tablica 3.8: ANCOVA za podatke iz tablice 3.7



Primijetimo da je  $p$ -vrijednost razlika u nagibu jednaka 0.186 te na razini značajnosti od 25% odbacujemo hipotezu o jednakosti nagiba. Pogledajmo sada drugi skup podataka iz [3]:

Seriya	Vrijeme(godina)	Sadržaj(%)	Seriya	Vrijeme(godina)	Sadržaj(%)
1	0,014	102,1	4	0,066	102,5
	0,280	103,5		0,343	102,4
	0,514	102,6		0,533	101,6
	0,769	101,9		0,802	102,6
	1,074	102,4		1,538	102,6
	1,533	102,4	5	0,011	102,5
	2,030	101,6		0,310	101,4
	3,071	101,8		0,624	100,3
2	4,049	102,6		1,063	100,2
	0,022	103,6	6	0,011	100,8
	0,118	104,0		0,310	99,9
	0,272	103,4		0,624	97,4
	0,566	103,8		1,063	100,1
	1,165	101,7	3	0,025	101,3
	2,022	103,4		0,275	100,6
3	0,025	101,3		0,547	100,2
	0,275	100,6		0,797	101,0
	0,547	100,2		1,041	101,8
	0,797	101,0		2,058	98,0
	1,041	101,8		2,519	100,1
	2,058	98,0			
	2,519	100,1			

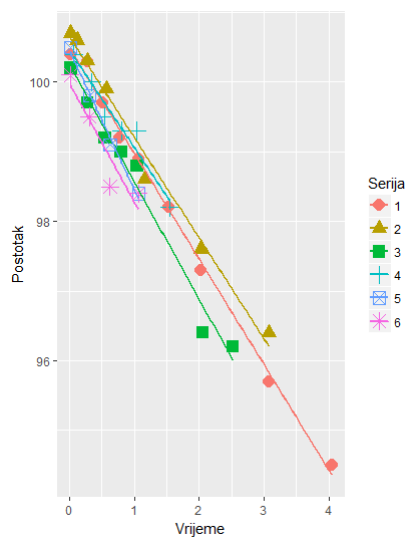
Tablica 3.9: Drugi skup podataka iz [3]

Analogno, pripadna ANCOVA tablica za podatke navedene u tablici 3.9 izgleda ovako:

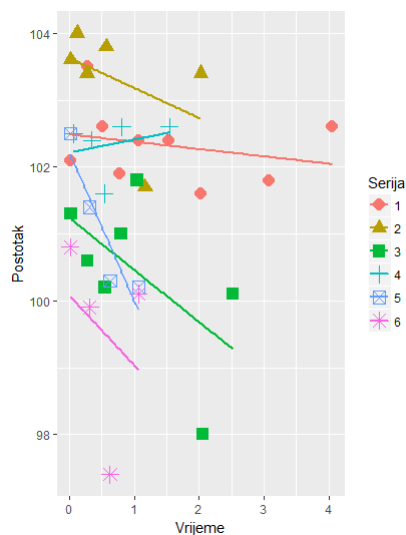
Izvor varijabilnosti	Stupanj slobode	Suma kvadrata	F	p
Odsječak	5	54.552	13.729	2.878e-06
Vrijeme	1	2.979	3.748	0.065
Razlika u nagibu	5	4.604	1.159	0.359
Greška	23	18.278		

Tablica 3.10: ANCOVA za podatke iz tablice 3.9

Primijetimo da je  $p$ -vrijednost razlika u nagibu jednaka 0.359 te na razini značajnosti od 25% ne možemo odbaciti hipotezu o jednakosti nagiba. Prikažimo prosječne pravce za svaku seriju zasebno, dobivene na podacima 3.7 i 3.9:



Slika 3.4: Prosječni pravci za podatke iz tablice 3.7



Slika 3.5: Prosječni pravci za podatke iz tablice 3.9

Kao što vidimo na grafovima 3.1 i 3.2, varijabilnost skupa podataka 3.7 je puno manja od skupa podataka 3.9. Međutim, držeći se smjernica danih u FDA regulativi i koristeći razinu značajnosti od 25%, test jednakosti nagiba zadovoljen je za 3.9. Ruberg i Stegeman ukazuju da dobro dizajnirane i pažljivo realizirane studije generiraju pouzdane skupove podataka s malom varijabilnošću. Međutim, preciznost izračuna procjenitelja za takve skupove podataka nije iskorištena u procjeni sveukupnog roka trajanja. U manje varijabilnom skupu podataka detektiraju se razlike koje nemaju veliki značaj za nagib, ali su statistički značajne zbog izbora razine značajnosti od 25%.

### 3.3 Implementacija u R-u

Prikažimo kod napisan u programskom jeziku R koji kao ulaz prima rezultat studije stabilnosti i kao rezultat vraća ANCOVA tablicu kao u 2.1:

```
ancova<-function(podaci)
{
  #=====
  K=max(podaci$Serija)
  N=length(podaci$Postotak)
  #=====

  #izračunajmo sume kao u (2.4)
  #=====
  S_xx_i<-numeric(K)
  for (i in 1:K)
  {
    pom<-as.vector(t(filter(podaci,Serija==i)%>%select(Vrijeme)))
    S_xx_i[i]<-crossprod(pom-mean(pom),pom-mean(pom))
  }

  S_yy_i<-numeric(K)
  for (i in 1:K)
  {
    pom<-as.vector(t(filter(podaci,Serija==i)%>%select(Postotak)))
    S_yy_i[i]<-crossprod(pom-mean(pom),pom-mean(pom))
  }

  S_xy_i<-numeric(K)
  for (i in 1:K)
```

```

{
pom_p<-as.vector(t(filter(podaci,Seri ja==i)%>%select(Postotak)))
pom_v<-as.vector(t(filter(podaci,Seri ja==i)%>%select(Vrijeme)))
S_xy_i[i]<-sum((pom_p-mean(pom_p))*((pom_v-mean(pom_v))))
}
#=====
#izračunajmo sume kao u (2.5)
S_xx_W=sum(S_xx_i)
S_yy_W=sum(S_yy_i)
S_xy_W=sum(S_xy_i)

#=====
#izračunajmo sume kao u (2.15)
#=====
pom_v=as.vector(t(select(podaci,Vrijeme)))
S_xx_T<-crossprod(pom_v-mean(pom_v),pom_v-mean(pom_v))

pom_p=as.vector(t(select(podaci,Postotak)))
S_yy_T<-crossprod(pom_p-mean(pom_p),pom_p-mean(pom_p))

S_xy_T<-sum((pom_p-mean(pom_p))*((pom_v-mean(pom_v))))

#=====
SSE_i=S_yy_i-((S_xy_i)^2/S_xx_i)
SSE=sum(SSE_i)

SSE_W=S_yy_W-((S_xy_W)^2/S_xx_W)
SS_beta=SSE_W-SSE

SST=as.numeric(S_yy_T)-((as.numeric(S_xy_T))^2/as.numeric(S_xx_T))
SSB=SST-SSE_W

SSS=(S_xy_W)^2/S_xx_W

#=====

MSE=SSE/(N-2*K)
MSB=SSB/(K-1)
MSS=SSS

```

```

MS_beta=SS_beta/(K-1)

#=====

F_alpha=MSB/MSE
p_alpha=pf(q=F_alpha, df1=K-1, df2=N-2*K, lower.tail=FALSE)
F_s=MSS/MSE
p_s=pf(q=F_s, df1=1, df2=N-2*K, lower.tail=FALSE)
F_beta=MS_beta/MSE
p_beta=pf(q=F_beta, df1=K-1, df2=N-2*K, lower.tail=FALSE)

#=====
#konstruiranje tablice
output=data.frame(df=numeric(4),sum_of_sq=numeric(4), F_stat=numeric(4),
                  p=numeric(4) ,stringsAsFactors = F)

output$df=c(K-1,1,K-1,N-2*K)
output$sum_of_sq=c(SSB,SSS,SS_beta,SSE)
output$F_stat=c(F_alpha,F_s,F_beta,NA)
output$p=c(p_alpha,p_s,p_beta,NA)

return(output)
}

```

Sada prikazimo kod koji kao ulaz prima rezultat studije stabilnosti i zadanu specifikacijsku granicu  $\eta$  te vraća rok trajanja izračunat za svaku seriju zasebno (uz pretpostavku da su odsječak i nagib različiti) i ukupni rok trajanja za sve podatke (uz pretpostavku da su nagib i odsječak jednaki):

```

shelf_life<-function(podaci,eta)
{

k=max(podaci$Serija)

output=data.frame(Serija=numeric(0),Shelf_life=numeric(0),
                  stringsAsFactors = F)
for (i in 1:(k+1))
{
  if (i==k+1) batch=podaci
  else batch<-filter(podaci,Serija==i)
}
}

```

```

Vrijeme=batch$Vrijeme
Postotak=batch$Postotak

x_potez<-mean(Vrijeme)
y_potez<-mean(Postotak)
n=dim(batch)[1]

S_xx<-as.numeric(crossprod(Vrijeme-x_potez,Vrijeme-x_potez))
S_yy<-as.numeric(crossprod(Postotak-y_potez,Postotak-y_potez))
S_xy<-sum((Vrijeme-x_potez)*(Postotak-y_potez))

beta_kapa<-S_xy/S_xx
alpha_kapa<-y_potez-beta_kapa*x_potez
s_2<-(S_yy-beta_kapa*S_xy)/(n-2)

SE_beta_kapa=(s_2/S_xx)^(1/2)
SE_alpha_kapa=(s_2*(1/(n*S_xx)*sum(Vrijeme^2)))^(1/2)

#=====dodatni uvijet=====
T_alpha_kapa=(alpha_kapa-eta)/SE_alpha_kapa
p_ak=pt(T_alpha_kapa,n-2,lower.tail = FALSE)

T_beta_kapa=(beta_kapa)/SE_beta_kapa
p_bk=pt(T_beta_kapa,n-2,lower.tail = TRUE)
#=====

if (p_ak<0.05 & p_bk<0.05)
{
  t=qt(0.05,n-2,lower.tail = FALSE)

  #računamo korijene jednadžbe (2.17)
  z=numeric(3)
  z[1]=eta^2-2*eta*alpha_kapa+alpha_kapa^2-t^2*s_2*(1/n+x_potez^2/S_xx)
  z[2]=-2*eta*beta_kapa+2*alpha_kapa*beta_kapa+2*x_potez*((t^2*s_2)/S_xx)
  z[3]=beta_kapa^2-(t^2*s_2)/S_xx
  sl_pom=as.numeric(polyroot(z))
  sl=min(sl_pom[sl_pom>0])

```

```
    } else sl=NA

    redak=data.frame(Seriya=i,Shelf_life=sl)
    output=rbind(output,redak)
  }

  output$Serija=as.character(output$Serija)
  output$Serija[nrow(output)]="Sve"

  return(output)
}
```

# Bibliografija

- [1] Sanford Bolton i Charles Bon, *Pharmaceutical Statistics*, Dekker Media, 2004.
- [2] Shein Chung Chow, *Statistical Design and Analysis of Stability Studies*, Chapman and Hall/CRC Biostatistics Series, 2007.
- [3] Stephen J. Ruberg i James W. Stegeman, *Pooling Data for Stability Studies: Testing the Equality of Batch Degradation Slopes*, <http://www.jstor.org/stable/2532658>.
- [4] Željko Pauše, *Uvod u matematičku statistiku*, Školska knjiga, 1993.



# Sažetak

U ovom radu opisana je metoda analize rezultata studije stabilnosti u svrhu određivanja roka trajanja lijekova. Prema trenutnim regulatornim propisima, spomenuta analiza je obvezan dio dokumentacije prilikom registracije novog lijeka na tržištu. Definiran je model koji opisuje degradaciju aktivnog sastojka lijeka s obzirom na vrijeme. Opisana je metoda analize kovarijance kojom ispitujeemo možemo li više nezavisnih serija lijeka smatrati jednakima te im pridružiti jedinstveni rok trajanja. Nakon toga opisan je način na koji određujemo rok trajanja pojedinog lijeka. Rok trajanja je sjecište pouzdanog intervala oko očekivanog pravca degradacije sa specifikacijskom granicom definiranom u postojećim regulativama. Kako bi se dobio što bolji uvid u opisane metode, dana je implementacija u programskom jeziku R te je opisana primjena na nekoliko primjera. Za kraj je dan primjer koji opisuje nedostatak navedene metode u smislu izbora razine značajnosti prilikom ispitivanja sličnosti više serija lijekova.

# Summary

This thesis is concerned with stability studies used to determine shelf-life of pharmaceutical products, in general, and drugs, in particular. Stability analysis is a mandatory part of an application procedure for a new drug product. We define a model for the active ingredient degradation of a drug product over time. Analysis of covariance method is described, to determine whether a single expiry date is suitable for several independent batches of a same substance. After that, we describe how to determine a unique shelf-life, following FDA guidelines. For a better understanding, we implemented described methods with the statistical programming tool R. At the end, we gave an example describing concerns about suggested level of significance when performing the test for batch similarity.

# Životopis

Rođen sam 26.12.1992. godine u Zagrebu, a odrastao sam u Zaboku gdje sam pohađao osnovnu školu do 2007. godine. Te godine, u istom gradu, upisao sam matematičku gimnaziju Antuna Gustava Matoša koju sam završio 2011. godine. Iste godine započeo sam studij matematike na Prirodoslovno-matematičkom fakultetu u Zagrebu, a 2014. godine upisao sam Diplomski studij Matematička statistika. Tečno govorim engleski jezik i dobro se snalazim u programskim jezicima kao što su R, SAS, MATLAB, C i C++. Od ove godine počeo sam raditi za tvrtku Cantab Analytica kao Senior Analyst. Bavim se analizom podataka i strojnim učenjem. U slobodno vrijeme volim gledati filmove i serije te zaigrati košarku i nogomet.